PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : A01H 5/00, C07H 21/04, C12N 15/63, 15/82

A1

(11) International Publication Number: **WO 98/44781**

(43) International Publication Date: 15 October 1998 (15.10.98)

(21) International Application Number: PCT/US98/06761

(22) International Filing Date: 3 April 1998 (03.04.98)

(30) Priority Data:
60/042,926 4 April 1997 (04.04.97) US

(71) Applicant *(for all designated States except US)*: PURDUE RESEARCH FOUNDATION [US/US]; Room 307, 1021 Hovde Hall, West Lafayette, IN 47907–1021 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: HODGES, Thomas, K. [US/US]; 6038 West 75 N., West Lafayette, IN 47906 (US). LYZNIK, Leszek, A. [PL/US]; 5960 N.W. 62nd Avenue, Johnston, IA 50131 (US).

(74) Agent: LAMMERT, Steven, R.; Barnes & Thornburg, 11 South Meridian Street, Indianapolis, IN 46204 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: REGULATORY ELEMENT FOR EXPRESSING GENES IN PLANTS

(57) Abstract

The present invention is directed to a novel promoter sequence for expressing genes in eukaryotic cells and the construction of expression vectors comprising that promoter sequence.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

REGULATORY ELEMENT FOR EXPRESSING GENES IN PLANTS

Field of the Invention

The present invention is directed to nucleic acid sequences that control the expression of genes in eukaryotic cells. More particularly, the invention is directed to a gene promoter that confers a high level of expression to genes that are operably linked to the promoter.

Background and Summary of the Invention

The present invention relates to a novel regulatory element which confers a high level of expression in plant cells to genes that are operably linked to the regulatory element. The ability to control the level of gene expression in plants is important for many applications of genetic transformation procedures including those directed to crop improvement.

In eukaryotic organisms, multi-level regulatory systems exist to control gene expression. The transcription process is an integral part of such systems and is involved in synthesis of mRNA molecules. The efficiency of transcription is mostly determined by a region of DNA called the promoter. The promoter consists of gene sequences upstream of the site of transcription initiation. The components of the promoter region include the "TATA" box and often a "CAAT" box. In addition, many other regulatory elements that affect transcription may be present in the promoter sequences. The coordinated action of cellular proteins (transcription factors) interacting with promoter sequences determines the specificity of a particular promoter and its effectiveness. Since most eukaryotic genes are stringently regulated, there is a limited availability of promoters with constitutive, strong expression.

The present invention describes the isolation and purification of a DNA sequence that expresses operably linked genes to high levels in plant cells. The promoter sequence described in the present invention expresses genes at a level equal to or higher than that obtained from one of the strongest presently available promoters - the 35S cauliflower mosaic virus promoter. Such promoters are needed to direct a high level of protein expression in transgenic plants. The strong promoter of the present invention is used to construct expression vectors for expressing genes in plant

cells. In one embodiment, a plant expression vector is provided that comprises the regulatory element of SEQ ID NO: 2 operably linked to a non-natively associated gene, and this vector is used to produce transgenic plants.

Brief Description of the Drawings

Fig. 1 Represents a restriction map of the 4.8 kb *Hind*III *Arabidopsis* genomic fragment that hybridizes to the *Bgl*I RTS-1 gene fragment.

Fig. 2 Expression of *gus*A in *Arabidopsis* protoplasts when *gus*A is operably linked to: the 35S cauliflower mosaic virus promoter (p35GUS), the promoter of SEQ ID NO: 2 (pUN-GUS), the promoter of SEQ ID NO: 3 (pASR-GUS) or lacking a promoter (DNA-).

Detailed Description of the Invention

Definitions

Unless specified otherwise, any reference to DNA, a DNA sequence, promoter, or regulatory sequence is a reference to a double stranded DNA sequence. A promoter is a DNA sequence that directs the transcription of a structural gene. Typically, a promoter is located in the 5' region of a gene, proximal to the transcription start site of a structural gene. If a promoter is an inducible promoter, then the rate of transcription increases in response to an inducing agent. In contrast, if the promoter is a constitutive promoter, then the rate of transcription is not regulated by an inducing agent.

An enhancer is a DNA regulatory element that can increase the efficiency of transcription, regardless of the distance or orientation of the enhancer relative to the start site of transcription.

The term "expression" refers to the biosynthesis of a gene product. For example, in the case of a structural gene, expression involves the transcription of the structural gene into messenger RNA and the translation of messenger RNA into one or more polypeptides.

An expression vector is a DNA molecule comprising the regulatory elements necessary for transcription of a gene in a host cell. Typically the gene is placed under the control of certain regulatory elements including constitutive or

inducible promoters, tissue-specific regulatory elements, and enhancer elements. Such a gene is said to be "operably linked to" the regulatory elements when the regulating element controls the expression of the gene. Expression vectors typically include eukaryotic and/or bacterial selectable markers that allow for selection of cells containing the expression vector.

An exogenous DNA sequence refers to a DNA sequence that has been introduced into a host cell from an external source. A transgenic plant is a plant having one or more plant cells that contain an exogenous DNA sequence. The term stably transformed refers to a transformed cell or plant that is capable of transmitting an exogenous DNA sequence to its progeny. Typically a stably transformed host has the exogenous DNA sequence integrated into its genome.

A core promoter contains the essential nucleotide sequences for promoter function, including the TATA box and start of transcription. By this definition, a core promoter may or may not have detectible activity in the absence of specific sequences (regulatory elements) that may enhance the activity of the core promoter or confer tissue specific activity.

A visible marker is defined herein as including any gene that encodes a product that produces a phenotypic trait to the host cell or organism.

A selectable marker is defined herein as including any nucleic acid sequence or gene product that can be selected for after introduction into a cell. The selectable marker facilitates the identification of transformants.

A polylinker is a DNA sequence that contains multiple endonuclease restriction enzyme identification sequences in close proximity of one another.

The present invention is directed to a substantially purified genomic DNA sequence isolated from *Arabidopsis thaliana* (SEQ ID NO: 1). The genomic DNA encodes for two proteins (ASR-2 and ORF 3) and contains a dual promoter region located between those two genes that drives the expression of both genes (see Fig. 1).

The genomic region containing the coding DNA sequence for ASR-2, located between nucleotides 945 to 3694 of SEQ ID NO: 1, encompasses sequences that are homologous to a human pre-mRNA splicing factor ASF/SF2 and the *Arabidopsis* SR1 gene. The alignment of the ASR-2 genomic DNA sequences with

the SR1 cDNA sequence indicated the presence of eleven putative exons in the ASR-2 gene, and the deduced amino acid sequence has 82% identity (92% similarity) with the deduced amino acid sequence of SR1. The sequence identity of ASR-2 with the human splicing factor SF2 was 62% as compared to 59% identity between the SR1 and SF2 genes. The ASR-2 gene also appears to have an identical structural organization of RNA-binding domains, the glycine spacer, and the SR domain as is observed in the SR1 and SR2 genes. The ASR-2 coding sequence also includes a highly charged PSK domain at the C-terminal end similar to the SR1 gene but absent in the ASF/SF-2 gene coding sequences.

The regulatory elements controlling the expression of the ASR-2 gene are contained within the 530 nucleotide region shown as SEQ ID NO: 3. The expression of the ASR-2 gene was analyzed by reverse transcription PCR in different parts of *Arabidopsis* plant. The ASR-2 gene was found to be expressed in all plant parts investigated including the leaves, stems, siliques, and roots. Similar levels of expression were observed in different plant organs. The experiment revealed the presence of more than one transcript hybridizing to the ASR-1 probe (the 2.4 kb *Eco*RI fragment of ASR-1 genomic clone), and could represent splice variants of ASR-2 transcripts.

The 4.8 kb HindIII genomic fragment encodes for another gene, ORF 3 that is located on the complimentary strand relative to the sequence encoding the ASR-2 gene (see Fig. 1), between nucleotides 4217 and 4917 of SEQ ID NO: 1. A 530 bp region is located between the ASR-2 and ORF 3 genes (at position 3691-4220 of SEQ ID NO: 1) and that 530 bp region functions as a dual promoter for expressing both ASR-2 and ORF 3. The sequence of the DNA region that contains the regulatory elements for expressing ORF 3 is shown as SEQ ID NO: 2.

SEQ ID NO: 2 and SEQ ID NO: 3 are inverse compliments of each other, and accordingly a double stranded DNA sequences that contains SEQ ID NO: 2 will also contain SEQ ID NO: 3. As used herein with reference to double stranded DNA sequences, SEQ ID NO: 2 and SEQ ID NO: 3 will designate the orientation of the 530 bp region in DNA constructs. If the 530 bp region is ligated to a gene through its 3' end (as shown in SEQ ID NO: 1), the sequence will be referred to as SEQ ID NO: 2, and if the 530 bp region is ligated to a gene through its 5' end, the sequence

will be referred to as SEQ ID NO: 3. For example, a gene operably linked to a promoter comprising the sequence of SEQ ID NO: 2 designates that the promoter is operably linked to that gene in the orientation naturally expresses the ORF 3 gene.

The 530 bp region located between the ASR-2 and the ORF 3 genes promoter region contains sequences that are known to bind proteins that are involved in the transcriptional process and can function in either direction. Accordingly, this sequence can be used in either orientation as a promoter for expressing genes in eukaryotic cells and more particularly in plant cells. The present invention is directed to a substantially pure DNA sequence comprising the sequence of SEQ ID NO: 2, and the use of such a sequence to express exogenous genes in plants.

In accordance with one embodiment, a recombinant expression vector is prepared comprising a promoter having a consecutive 20 base pair sequence identical to the sequence of SEQ ID NO: 2 or SEQ ID NO: 3. Typically the expression vector will also include a polylinker region located adjacent to the promoter such that upon insertion of a gene sequence into the polylinker, the gene will be operably linked to the promoter. In one embodiment the promoter utilized is the DNA sequence of SEQ ID NO: 2. The expression vector typically includes a eukaryotic selectable marker gene or a visible marker gene to allow identification of plant cells transformed with the exogenous DNA sequence. In one embodiment the expression vector further includes a prokaryotic selectable marker gene and a prokaryotic origin of replication that allow for the transformation and reproduction of the expression vector in prokaryotes.

In accordance with the present invention, a DNA construct comprising the regulatory element of SEQ ID NO: 2, a core promoter and a gene operably linked to the core promoter is used to transform a plant cell, using procedures known to those familiar with the art. Such transformation procedures include but are not limited to microinjection, microprojectile bombardment, electroporation, calcium chloride permeabilization, polyethylene glycol permeabilization, protoplast fusion or bacterial mediated mechanisms such as *Agrobacterium tumefaciens* or *Agrobacterium rhizogenes*.

Transformed cells (those containing the DNA inserted into the host cell's DNA) are selected from untransformed cells through the use of a selectable

marker included as part of the introduced DNA sequences. Transformed cells/plant entities can also be identified by the expression of a visible marker included as part of the introduced DNA sequences. Visible markers include genes that impart a visible phenotypic trait such as seed color (i.e., yellow, purple or white genes) or shape (i.e., shrunken or plump genes). Selectable markers include genes that provide antibiotic resistance or herbicide resistance. Cells containing selectable marker genes are capable of surviving in the presence of antibiotic or herbicide concentrations that kill untransformed cells. Examples of selectable marker genes include the *bar* gene which provides resistance to the herbicide Basta, the *npt*II gene which confers kanamycin resistance, and the *hpt* gene which confers hygromycin resistance. An entire plant can be generated from a single transformed plant cell through cell culturing techniques known to those skilled in the art.

In one embodiment a transgenic plant entity is provided wherein the plant entity consists essentially of a plant cell, seed or plant produced from the *in vitro* introduction of an exogenous nucleic acid sequence into a plant cell, wherein the exogenous nucleic acid sequence encodes a gene whose expression is controlled by the regulatory elements of SEQ ID NO: 2. More particularly, the transgenic plant is generated by transforming a plant cell with a DNA vector comprising a promoter, having a consecutive 20 base pair sequence identical to the sequence of SEQ ID NO: 2 operably linked to a gene. In one embodiment, the DNA vector used to transform the plant cell comprises the 520 bp sequence of SEQ ID NO: 2 operably linked to a gene. The gene may encode for any product that is beneficial to the plant (for example, gene products that directly or indirectly provide herbicide resistance, insecticidal resistance, fungal resistance or act as growth regulators) or may encode for pharmaceutical or polymer components that are subsequently purified from plant material for commercial use. The exogenous nucleic acid sequences used to produce the transgenic plant typically also include a selectable marker gene or a visible marker gene to allow identification of the cells transformed with the exogenous DNA sequence. In accordance with one embodiment, a plant expression vector comprising a regulatory element operably linked to a non-natively associated gene is used to produce a transgenic plant, wherein the regulatory element is selected from the sequence of SEQ ID NO:2.

The regulatory element of SEQ ID NO:2 has been demonstrated to be highly efficient in transcribing genes in *Arabidopsis* cells (see Example 2 for details). As shown in Fig. 2 the regulatory element of SEQ ID NO: 2 ligated to the *gusA* coding sequence induced GUS activity at the level of 1.72 nmol MU/hr/1,000 protoplasts, whereas the 35S Cauliflower Mosaic Virus promotor when operably linked to the *gusA* coding sequence produced GUS activity at 1 nmol MU/hr/1,000 protoplasts. The regulatory element of SEQ ID NO: 3 ligated to the *gusA* coding sequence exhibited low level of GUS activity in *Arabidopsis*. Accordingly the 530 bp region, as shown in SEQ ID NO: 2, functions as a strong promoter when operably linked to an exogenous gene in the orientation that naturally expresses the ORF 3 gene in *Arabidopsis*.

Example 1

Isolation of the Genomic Fragment Encoding SEQ ID NO: 1

A genomic library of *Arabidopsis thaliana* ecotype RDL (prepared by ligation of *Hind*III partially digested genomic DNA fragments, ranging between 8-23kb, into the *Hind*III site of the binary cosmid pBIC20) was screened with a *Bgl*I fragment of the rice anther-specific cDNA clone RTS-1 (SEQ ID NO: 4) to isolate DNA fragments containing homologous sequences.

The RTS-1 cDNA clone is a tapetum specific gene that encodes an alanine-rich protein that is expressed in tapetum cells of rice anthers. The gene is more fully described in PCT application serial no. PCT/US96/16418, published on April 17, 1997 (publication no. WO97/13401), the disclosure of which is expressly incorporated herein.

Library screening was performed in large Petri dishes (20x20 cm) containing approximately 20,000 recombinant colonies of *E. coli* NM554 cells. Such density should represent about three *Arabidopsis* genome equivalents. The recombinant colonies were lifted on Hybond-N hybridization transfer membranes (Amersham) and membrane-bound DNA (UV irradiation) was probed with the *Bgl*I cDNA fragment of the RTS-1 gene (SEQ ID NO: 4). Membranes were prehybridized at 50°C for 1 hr in pre-hybridization solution containing 5xSSPE, 5x Denhardt's solution, 0.5% SDS, and 0.2 mg/ml denatured salmon-sperm DNA. Hybridization was

overnight at 50°C. The filters were washed twice at 50°C in 3 x SSC solution for 15 min, once at 50°C for 15 min in 1 x SSC solution, and in 0.2 x SSC solution at 50°C for 15 min followed by 30 min incubation at room temperature. Washed filters were wrapped in SaranWrap and autoradiography was carried out overnight.

Twenty-three independent clones hybridizing to the probe were identified and selected for endonuclease restriction analysis and Southern blotting. Most of the clones gave rise to multiple signals of varying intensity upon probing with the *Bgl*I cDNA fragment of the RTS-1 gene (SEQ ID NO: 4). The initial endonuclease restriction and Southern blot analysis identified genomic clone #2 as having a 4.8 kb *Hind*III fragment that hybridizes to the RTS-1 probe. This clone was selected for further detailed analysis. When the 4.8 kb *Hind*III fragment was subsequently restricted into the two 2.2 kb and 2.4 kb *Eco*RI fragments, both fragments hybridized to the probe indicating two independent probe binding sites.

The 4.8 kb *Hind*III DNA fragment was sequenced using standard techniques. For sequence analysis, the 2.2 and 2.4 *Eco*RI fragments internal to the 4.8 kb *Hind*III fragment (See Fig. 1) were subcloned into the pBluescript KS +/-vector. In addition, DNA fragments generated by digestion of the *Eco*RI fragments with *Xba*I were subcloned to facilitate the sequencing process. The sequence of the cross-hatched region shown in Fig. 1 is shown as SEQ ID NO: 1

A simple homology search for sequences similar to the RTS-1 probe resulted in the identification of three possible binding sites within the 4.8 kb *Hind*III fragment (indicated as boxes above the cross-hatched region of Fig. 1). The matching percentage was in the range of 35-39% over the 190 bp probe fragment. Experimental results on restriction fragment hybridization to the RTS-1 probe were in agreement with predicted positions of the probe binding sites. The longest open reading frame is located in one region of probe binding and it was selected for further analysis.

A genomic fragment from position 825 to 3694 nucleotide SEQ ID NO: 1 was identified as containing sequences homologous to the human splicing factor ASF/SF2 and the *Arabidopsis* SR1 gene, and that region was designated as the ASR-2 region (See Fig. 1). The alignment of genomic DNA sequences of this gene with the SR1 cDNA sequences indicated the presence of eleven putative exons with 82% identity (92% similarity) of the deduced amino acid sequences. The sequence identity

to the human splicing factor SF2 was 62% as compared to 59% identity between the SR1 and SF2 genes. An identical structural organization of RNA-binding domains, the glycine spacer, and the SR domain was observed among all three genes. The ASR-2 coding sequence also included a highly charged PSK domain at the C-terminal end similar to the SR1 gene but absent in the ASF/SF-2 gene coding sequences.

The presence of ASR-2 transcripts was analyzed by reverse transcription PCR in different parts of *Arabidopsis* plant. Total RNA was isolated from various *Arabidopsis* organs and reverse transcription of 5 μg of the total RNA, treated with RNase-free DNase, was performed with MuMLV-reverse transcriptase (400 units) and oligo $dT_{18\text{-}22}$ primer (4 μg) for 1 h at 42°C followed by 5 min at 95°C. Following incubation, the reaction mixture was treated with RNAse H (8 units) for 20 min at 37°C. Five μl of the reverse transcription reaction was then amplified with Taq-polymerase (Perkin-Elmer Cetus) using primers that recognize the first exon in the RNA recognition domain and the SR domain of ASR-2. Primer sequences were selected that were specific to the ASR-2 domains but not to the SR1 homologous domains.

The ASR-2 gene was found to be expressed in all plant parts investigated including the leaves, stems, siliques, and roots. Similar levels of expression were observed in different plant organs. The experiment revealed the presence of more than one transcript hybridizing to the ASR-2 probe (the 2.4 kb EcoRI fragment of ASR-2 genomic clone). Shorter transcripts were identified in RT-PCR reaction products than expected and could represent splice variants of ASR-2 transcripts. Sequencing of the amplified major RT-PCR product confirmed all predicted intron-exon junction sites except the splicing sites (5' as well as 3') of the intron #7. Such transcripts contain the SR domain message but they cannot be translated into the full length protein because splicing of the intron #7 generates a frame shift mutation leading to a stop codon just after the splice site.

Example 2

Expression of gus A using the promoters of SEQ ID NO: 2 and SEQ ID NO: 3

Arabidopsis protoplasts were transformed with bacterial vectors containing the 5' untranslated ASR-2 DNA sequence connected to the coding

sequences of the bacterial b-glucuronidase gene. The coding sequences were ligated to the 3' end and to the 5' end of the promoter sequence and the respective constructs were designated as pUN-GUS [having the promoter orientated in the direction that normally transcribes the ORF-3 gene (i.e., SEQ ID NO: 2) and operably linked to the *gusA* gene] and pASR-GUS [having the promoter orientated in the direction that normally transcribes the ASR-2 gene (i.e., SEQ ID NO: 3) and operably linked to the *gusA* gene]. The sequences of the junction site between the promoter sequence and the *gusA* coding sequence are disclosed as SEQ ID NO: 5 for pASR-GUS and SEQ ID NO: 6 for pUN-GUS wherein the ATG start codon is located at nucleotide 6 and the coding region of the *gus*A is located at nucleotide 30.

The unique DNA promoter sequence discovered and claimed in the present invention is located between the nucleotides at the position 3690 through 4221, of SEQ ID NO: 1. The sequence is 530 nucleotides in length and is presented as SEQ ID NO: 2 and SEQ ID NO: 3. The 530 bp sequence contains numerous transcription factor binding sites including two "TATA" boxes at positions 141, 316 and "CAAT" boxes located at 223, 386, 448, and 486 and one *zeste* element (GTGAGTG) at 264 of SEQ ID NO: 2.

The activity of the claimed sequence driving the expression of a foreign gene, *gus*A, in plant cells was compared to the activity of the 35S cauliflower mosaic virus promotor in *Arabidopsis* protoplasts. Four expression vectors p35SGUS (having the 35S cauliflower mosaic virus promotor operably linked to the *gus*A gene), pUN-GUS (as described above), and pASR-GUS (as described above) and a control vector lacking a promoter operably linked to the *gus*A gene (DNA-), were introduced into the protoplasts by a PEG-mediated transformation procedure. One day after transformation, the GUS activity was determined. The 35S CaMV promotor controlling the *gusA* sequence produced GUS activity at 1 nmol MU/hr/1,000 protoplasts, while the claimed sequence ligated to the *gusA* coding through its 3' end (SEQ ID NO: 2) induced GUS activity at the level of 1.72 nmol MU/hr/1,000 protoplasts (see Fig. 2). The claimed sequence ligated to the *gusA* coding sequence in the opposite orientation (SEQ ID NO: 3) exhibited a low level of GUS activity in *Arabidopsis*.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Purdue Research Foundation,
Hodges, Thomas K.
Lysnik, Leszek A

(ii) TITLE OF INVENTION: Regulatory Element For Expressing Genes In Plants

(iii) NUMBER OF SEQUENCES: 6

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Barnes & Thornburg
(B) STREET: 11 S. Meridian
(C) CITY: Indianapolis
(D) STATE: Indiana
(E) COUNTRY: USA
(F) ZIP: 46204

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Breen, John P.
(B) REGISTRATION NUMBER: 38,833
(C) REFERENCE/DOCKET NUMBER: 3220-29933

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (317) 231-7745
(B) TELEFAX: (317) 231-7433

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5285 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Arabidopsis thaliana

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GAATTCCAGC GTGGAAGAGA CCAGGACAAC AAACAGCGAG TTTGTATAAA GAAGCCCAAC 60

CACCGGGAGG AGTAAGAGAC GAATCCGCCG CGGTGGATGA AGAAGCGGAT GCGGCGGCGG 120

AGGGAGGAAG AGGGAGGAGA TCGGAGATGA CGGTGGGGAT AGCGGAGTAG CACTGAGTTT 180

TACAACAAAC GTTTCGAACA ATTGAGGAGA ATTTGGTGCA GACGCAAGCG AGACGAGCCC 240

AGTTTCTTGG ATCATCTTGG AGCTTGAAGA AGATGTTGAA GACTACGTCT TCTGGTATAC 300

AAGAGAAGAC AGATTCCGCC ATGGATCGCC TTCTTCTCTC TAGGCGGCTC CTTTCTTATC 360

CAAATTCACT TATACTGTTA TGGGTCCGGT ACGCGTAAAC CGGGAATAGT CTTAACTGTT 420

CTTAAGGTGG GTCACAGATT CACTAACACC CACACAAAGG CAAGTAAGTA ATGCGCAACA 480

GCTCTCGAAA ATGACATCGT ACGGACTGAA CTAAAATGTA AAGGGTCCGG GTATCAAAAT 540

GAGTTCAATG CACATGTCTT TTTAGGTTCA TTTATTGTGA ACGTTTTCAA AATTTTAATA 600

TCGAATTGTG AGCTTTTGAA TTAAGTTTGG TATTCGACAG TAATTTTTGA TAGTTCGTTT 660

TAAGCACTAA CTATATTAGC AAGTCATATA AATCAGCTGA GCTTAGCTCA TAAACTGATG 720

ATGACTGAGT ATATATCATT CCCATGTGCA AACCCAAGCT AATAAGAATG AAATACACAA 780

CTAGTTTTTC AACTTCTCAT ACATAAGAGA GATCATCTTT ATGAGAATCT TCCAACAGAC 840

CCAGCTCTTT TACCTGATAT GAGATGTTTC TTTACCTGTA AAACAGTTGG AAAGAGATTC 900

AGAGAAAGAG TGAGAATGTT CCAAGCAAGT TAAAAAGAGT GTGCATTACC GAGATGGACT 960

CCTGCTTCTG CTCTTGCTCC TACCACGGAT AGGGCTCAGC TGCTTGCTAG GGCTCTTACT 1020

TGCTTCCTTC TGAACCTGCA ACACACACAA TTCAATCCAG ATGATGAGAA TGTAATTAGC 1080

TAGACATAGA TCATGTCTTT TGGATAGTAT GGATTGAAAA CCGAAAATAT TGTGCTGATA 1140

GTATAGCGGA GTAAAAAGTG TATTAGAGAT AGATGACTTA GAGAGGGTAG AGGAGATCTT 1200

GACCTTGAAC GAGATCTGCA AAAGTCCGAG AAACAAATCC AGTTTTAAAA ATCCAATATT 1260

TCCGTGTTTA ATAGCTGTTT CACCAAGACC CTTGATGCCT CATTCCATGT CAGACATCAA 1320

AACACAATCA CAAGAAGCCA AAAAGGAAAA AACATATCTT CAGACTATTT TTAGCAGAAA 1380

ATCAAACCAA CACAAGCCTG CTTTTTGTTA AAACAGATGG TAGAGATGGA GAATATCAAA 1440

GCAAATCTAA TTTTATCAAA CCCTGTGTCC AGAACAGCAT CGGTTCCATG AGAATCCAGA 1500

TCGCTTGTCG ATTAGCAGGT AATGAAAATA TCTGTCCTAG CTCGATGGGT CCATTTTGAT 1560

GACACTATAT CAATGCGATC CAATGTCTCT CCACTGTTAC CCATTTAGCA GCAATGGATA 1620

TTATCAGAAA ACGAACTTGC CCATTTAAGA AAGAGCATAT ACCTTTGAGA CGAAACAGTT 1680

GTGTACAACC AACCAAGACA TTCCATGATC CAGATACTAT TTCCCATATT TTAGTTGATT 1740

GTATGTATAT CCATATCTAA GAAACAAAAC CATTCTCAAC ACTATAATTA TAAAAGACCA 1800

GACTTTCAAA GGAAATAAAT CGTGTACCCT TAACTAGAAG CAATCATCAT TCTCAATCAA 1860

CAACTCGGAG TCATCCCAGT GACATCTTTT AATGTGATGT CACCAAACTT CAAGGGAAGA 1920

GCTAAATAGA GCCTATGTTA TGTTTTATGT TGGATATTTT AGCATAAACA TTATAGAAGA 1980

AATGAAGCAT ACCCTCGTGG AGACAGTGAC CTCGACTTAG AGCGGGAGCG AGAGCGAGGA 2040

GATCTCGATG TAGATTTTGC AGGCGATCTA CGCAAAGATT TAGCCTTTGG ACTTCTGCTC 2100

TTGCTCCTGC TTCTGCTGCG GCTACGACTA CGGCTGGGAC TCCGTCCACG GCTGCGGCTC 2160

TTAGAATAGG ATCTTCCACG GCTGGGGCTC CTCGAATCCC TCCTTGAATC ATATTCTCTA 2220

ACCTGTAATG ACATAGGGAA ATGTTAAGTG CAATAAAAAC GAATGCAGCT CTCAGAATTC 2280

TTGTCTTTAA CATACCCGAA CATATTCATG AGAAAACGCA TTCCGAAACT CTGTGTCATC 2340

GAGCTTTTTT TATCTGGACA AAATAAAGAA TATAATCATG AGTAATCAAG GATGAACACA 2400

TTCTCAGCCA CAGTCCCACA CAGAAAATCA AATAATATGA AGAGGAAGAA AAACATCTCA 2460

CCGCATATTT CATGTCCTCG TAGCTGGTAT AATCTACAAT TCCAGTTGTA CCTGTAAATA 2520

AATAGTCCAC GATATAGATT TTTTAAAGCA TGCACAGACT AGTTAAAACA GGATTTAAAG 2580

GCAGAAAACC CAAACAGCTT TAGACACATC TCTATTCTTG GGTAAGACAT GAGGATTTAC 2640

CTCTACCATC ACGAAACACT TGAGAAAAAC AAACTTCTCC TCCTTTACGC ATGTGATCCT 2700

TCAAAACAAA GTGATATGTC AACATTCAGA AATCGTAGAA AATATAGGAA CGACAATGAG 2760

GAATCTGTCC ACAACTGTGT AATCACCTTG AGGTCTTGCC AGGACGCAGA TGAAGGCAAA 2820

CCTGACACTA CAACTGTATA GTGGAAAATC TTAATTTAGT GATTTCTCCT AAAACTTATG 2880

AATACACTAA AGCTAAACAT ATCATATGTA CCGCGGTACT CTGATCTCCT AGATGGTCCA 2940

CGTTCACGAC CACCACCGTC ACCACCACCA CGACCGCCAC GACCACGACC ACTATAACTA 3000

CCGCGTGCAT CATGTGATGA ACGCCTCCCA CCATGAGCTA GTTCCACCTG CAATGGCCAA 3060

CACACATAAA TTATGTTTGG CTACCAGTCA ACAATACAAA GTTTGTGTAA AAATTCTGAA 3120

ATTTGATGAT TAACAAACCC GTAAATGATG CCCATCAAAG TCATAACCAT CACGGCCATA 3180

AATTGCATCA TCAGCATCAC GAGCATCCTC AAACTAAATC ACATATATCA CAAAACATTA 3240

GTGGTAGTAT CTCCCAACAT TTGAAAACTC ATGAACACTC AACAACAACG AAGAGCCTAA 3300

CTGTAAAATC AACAACAAGC CTTTATAAAA CATGTGGTTG CATAAAAAAT CTGACCTCGA 3360

CGAATGCATA GCCTGGAGGC CTCGGCGGAA TCTTCAAATC GATTTGAACA ACAGGTCCAT 3420

ACTTCAAAAA AAATAAGGAA GAACAATTCT TAAGGAAACT TCTTCCAAAT AAAATCAGAA 3480

TCCAACAATT CCGAAGATGC ATTCAATTTC TTTCACATGC AAATCTTGTA AAGACATATT 3540

CATCACATAA CACAAAATTC GATCCTGAGT TCTGAGTTCT TAAATTAGGA GAAACGAGTA 3600

ATTTACCTTA CTGAACAAGT CTTCAACTTC TCTTTCACGG ATATCGCCGG GGAAGGTCCC 3660

GACGTAAATC GTTCTACTCG AACGGCTGCT CATTTATTTC TTTCCTATAC CAAAATCAAA 3720

ATTGAGATTC GAAACGTCAA TAGATCGAAA CAAAGAAGCG ATCACACACA AAAAAAACTC 3780

ATTGGATAAC GATTAACCTA AGGAAAACTA AAGAGGTTTG ATTGATCGTC TATATATGAA 3840

CTAAAATTCC AGTAACGATT CCGATCACCT GAGAGAAAAT TCCGATGGAA GAGAAGAAGA 3900

AAGGCGAAAA TTGAAATCTG ACTAGGGCTT TCGAATACCA TAGAGATCAT CACGTGAGTC 3960

ACGTGACCGA CCGGGTACGT ATTAAAATAC ATTGTGTCTT GACCGTATAA AATACATTTG 4020

ACCCGTTTTG CAACAAATCG TAATCTTCAA TCAAAAGCTC TTAAACCCAA AAGAACAATT 4080

CCAAATCTTC AATACTTGAT ATTTCTCAAA GAACTTGAAA ACAACACAGA TCCATTCCCA 4140

ATTTCAGATT CACTCAAAAA GGATTTTTCT TTTTTCATTT TCGCTTTTTG TGATCTGGAA 4200

AGTTGTTACC TTTAACAATG TCTCCGAAAC ATCTAGAGTC ATCACGAAGC TCTATTGAAT 4260

CATGCACTTC ACAGCTTCTC TCATGGCGAC CATTTCACCG CTCCAAAACC CTAGACTCAT 4320

CTGACCAACC ACCGCAGACC AATGGGTTTC ACTCCTTTAC TCCCAAACGC CCTTGCTTCT 4380

CCGATCGATC CACTTCTTTC ACCATCGAAG CTATGAGCCG TCTCTCACTC GCCGACGACG 4440

ACAATGGAGG GAAGACATTA TCAGCTTCCA ATTACAGCAA TAGAGGAAGT TTCAGGTTAG 4500

TAGCGAGGAA GCGGCGGCGG CGTAATTCGA GATCGGTGTC TGGTCGGAGT AGTGATCGGA 4560

GTGGGACTCG GAGATGTTGC TCCATTGGTG CTCATGGGAC TTGTTCGGAT TTGCCTTTCG 4620

CTGTTGGTAC AGATTCAAGT GGAGAGCTTT TTGGTGAAGC GAATTGGGCT TCTGATGTGA 4680

GTGAGGCGGC GAGGAATTCA CGGAGAGAGC GGCGAGATTC TGGTGGAGAG AAGGAAGCTT 4740

CTGGTGGATT TGGATTTGCT AATGGAGTTG ATCCAATGGG GAATGAATCT GGGTATGGGA 4800

GTGAGCCTGG TTACAGAGGT GATGCTGAGT TTGGCTATGG TGATGAATTT GATGATGAAG 4860

AAGAAGATGT CGAGCCATTG TTTTGGGGAG GTATTAAATT CAGAGACTTT TTATAGCAAT 4920

TGTGTTCCAT CTTGAGATTC GTGGTTTTTG CTATGAAGAT TTGGAGATTG ATCATCATTG 4980

ATTAGATTAA AGATGACAAC TTTAGTGTTA TTTCTTCTGA TGAAAATGAG TCTGATTTTG 5040

CTCTGCTTGT CTATTATGGC ATTGCCTCAT AGGAATTGTC AGAAAGTTGT CAAATTTTGA 5100

TATGTTTAGT GATTGGTGAG TGTTTTGGAT GGAATTGGGT TCTTATCATG TTAGGTCATT 5160

GTCTGAAATG GATATGTATG TACTTGGTAT TTTGATATGT TTAGTGATTG GTGAGTGTTT 5220

TGGATTTGGA GCAGATACAG ATTCCACAAT GGGGATGTCT GGTGAGACAA ATCTCAGATA 5280

GTAAA 5285

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 530 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

CATTTATTTC TTTCCTATAC CAAAATCAAA ATTGAGATTC GAAACGTCAA TAGATCGAAA 60

CAAAGAAGCG ATCACACACA AAAAAAACTC ATTGGATAAC GATTAACCTA AGGAAAACTA 120

AAGAGGTTTG ATTGATCGTC TATATATGAA CTAAAATTCC AGTAACGATT CCGATCACCT 180

GAGAGAAAAT TCCGATGGAA GAGAAGAAGA AAGGCGAAAA TTGAAATCTG ACTAGGGCTT 240

TCGAATACCA TAGAGATCAT CACGTGAGTC ACGTGACCGA CCGGGTACGT ATTAAAATAC 300

ATTGTGTCTT GACCGTATAA AATACATTTG ACCCGTTTTG CAACAAATCG TAATCTTCAA 360

TCAAAAGCTC TTAAACCCAA AAGAACAATT CCAAATCTTC AATACTTGAT ATTTCTCAAA 420

GAACTTGAAA ACAACACAGA TCCATTCCCA ATTTCAGATT CACTCAAAAA GGATTTTTCT 480

TTTTTCATTT TCGCTTTTTG TGATCTGGAA AGTTGTTACC TTTAACAATG 530

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 530 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CATTGTTAAA GGTAACAACT TTCCAGATCT CAAAAAGCGA AAATGAAAAA AGAAAAATCC 60

TTTTTGAGTG AATCTGTTAT TGGGAATGGA TCTGTGTTGT TTTCAAGTTC TTTGAGAAAT 120

ATCAAGTATT GAAGATTTGG AATTGTTCTT TTGGGTTTAA GAGCTTTTGA TTGAAGATTA 180

CGATTTGTTG CAAAACGGGT CAAATGTATT TTATACGGTC AAGACACAAT GTATTTTAAT 240

ACGTACCCGG TCGGTCACGT GACTCACGTG ATGATCTCTA TGGTATTCGA AAGCCCTAGT 300

CAGATTTCAA TTTTCGCCTT TCTTCTTCTC TTCCATCGGA ATTTTCTCTC AGGTGATCGG 360

AATCGTTACT GGAATTTTAG TTCATATATA GACGATCAAT CAAACCTCTT TAGTTTTCCT 420

TAGGTTAATC GTTATCCAAT GAGTTTTTTT TGTGTGTGAT CGCTTCTTTG TTTCGATCTA 480

TTGACGTTTC GAATCTCAAT TTTGATTTTG GTATAGGAAA GAAATAAATG 530

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 186 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Oryza sativa

(vii) IMMEDIATE SOURCE:
(B) CLONE: RTS-1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GAGCCGCCCA CCGATGACGG CGCGGTCCGG GTGGCGGCGG GGCTGACGAA GTGCGTGTCC 60

GGGTGCGGTA GCAAGGTGAC CTCCTGCTTG CTCGGCTGCT ACGGCGGCGG CGGCGGCGCC 120

GCCGCCGCCG CGACGGCGAT GCCGTTCTGC GTCATCGGCT GCACCAGCGA CGTCTTGTCC 180

TGCGCC 186

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 38 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

AATAAATGAG CCCGGGTGGT CAGTCCCTTA TGTTACGT 38

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 38 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

TAACAATGTC CCCGGGTGGT CAGTCCCTTA TGTTACGT 38

CLAIMS:

1. A substantially pure nucleic acid sequence comprising a sequence as set forth in SEQ ID NO: 2.

2. A DNA sequence comprising a 20 base pair nucleotide portion identical in sequence to a consecutive 20 base pair portion of the sequence as set forth in SEQ ID NO:2.

3. A recombinant expression vector comprising the DNA sequence of claim 2 operably linked to a gene.

4. The expression vector of claim 3, further comprising a gene encoding a eukaryotic selectable marker.

5. The expression vector of claim 4, further comprising nucleic acid sequences that enable replication of the expression vector in a bacterial host, and a gene encoding a bacterial selectable marker.

6. A plant entity consisting essentially of a plant cell, seed or plant produced from the *in vitro* introduction of an exogenous nucleic acid sequence of claim 3.

7. An expression vector comprising the DNA sequence of claim 2 and a polylinker sequence.

8. The expression vector of claim 7, further comprising a gene encoding a eukaryotic selectable marker.

9. The expression vector of claim 8, further comprising nucleic acid sequences that enable replication of the expression vector in a bacterial host, and a gene encoding a bacterial selectable marker.

10. A transgenic plant entity comprising plant cells transformed with a DNA sequence comprising a regulatory element operably linked to an exogenous gene, said regulatory element comprising a 20 base pair nucleotide portion identical in sequence to a consecutive 20 base pair portion of the sequence as set forth in Seq ID NO:2.

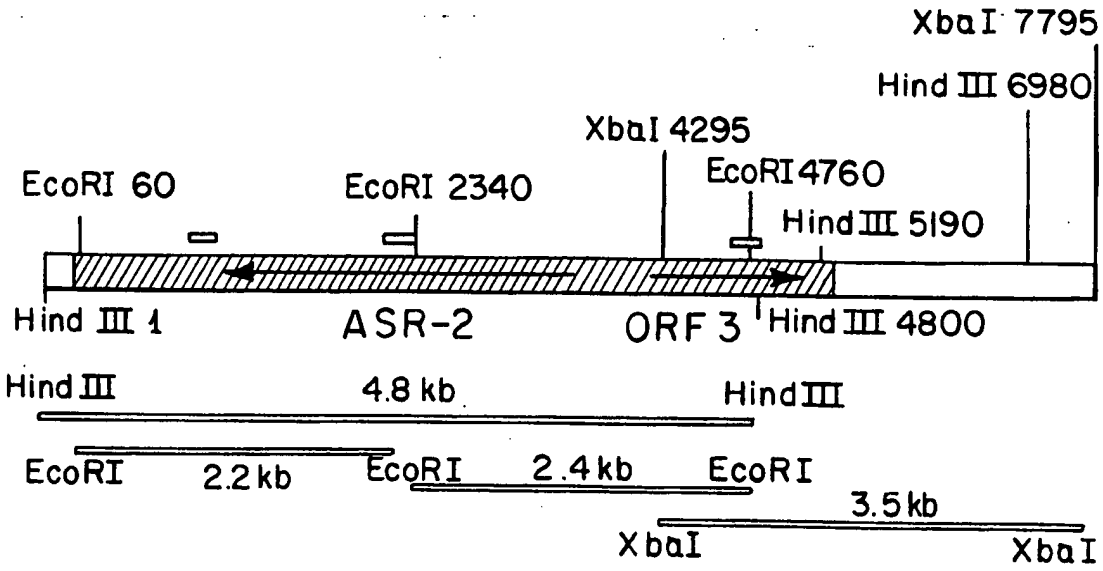
XbaI 7795
Hind III 6980
XbaI 4295
EcoRI 60
EcoRI 2340
EcoRI 4760
Hind III 5190
Hind III 1
ASR-2
ORF 3
Hind III 4800
Hind III
4.8 kb
Hind III
EcoRI
2.2 kb
EcoRI
2.4 kb
EcoRI
3.5 kb
XbaI
XbaI

FIG. 1

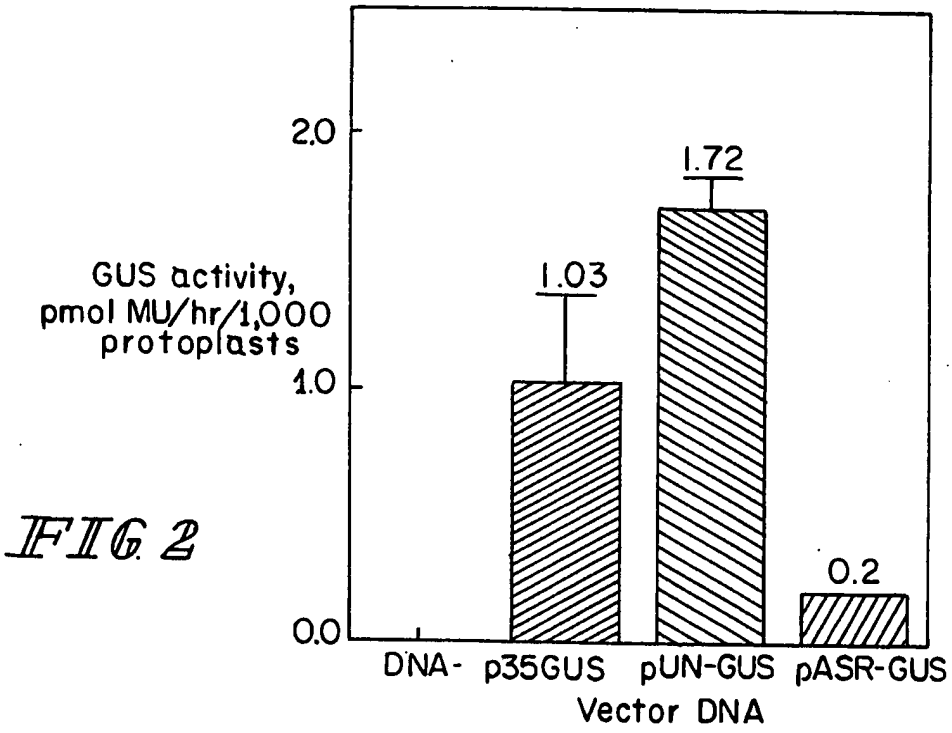
GUS activity,
pmol MU/hr/1,000
protoplasts
2.0
1.0
0.0
1.03
1.72
0.2
DNA-
p35GUS
pUN-GUS
pASR-GUS
Vector DNA

FIG 2

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/06761

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :A01H 5/00; C07H 21/04; C12N 15/63, 15/82
US CL :435/320.1; 536/24.1; 800/298

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/320.1; 536/24.1; 800/298

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	FISSCHER et al. Identification of Potential Regulatory Elements in the Far-Upstream Region of the Arabidopsis thaliana Plastocyanin Promoter. Plant Molecular Biology. November 1994, Vol. 26, No. 3, pages 873-886, see entire document.	1-10
A	DOLFERUS et al. Differential Interactions of Promoter Elements in Stress Responses of the Arabidopsis Adh gene. Plant Physiology. August 1994, Vol. 105, No. 4, pages 1075-1087, see entire document.	1-10
A	LAZAR et al. Identification of a Plant Serine-Arginine-Rich Protein Similar to the Mammalian Splicing Factor SF2/ASF. Proc. Natl. Acad. Sci. USA. August 1995, Vol. 92, pages 7672-7676, see entire document.	1-10

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 29 JULY 1998	Date of mailing of the international search report 18 AUG 1998
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer THANDA WAI Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/06761

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	LOPATO et al. Characterization of a Novel Arginine/Serine-Rich Splicing Factor in Arabidopsis. Plant Cell. December 1996, Vol. 8, No. 12, pages 2255-2264, see entire document.	1-10
A,P	SU et al. Arabidopsis ASF-SF2 Transcripts are Alternatively Spliced. Plant Physiology. July 1997, Vol. 114, No. 3, Supplement, page 246, see Abstract No. 1260.	1-10

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/06761

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

APS, MEDLINE, AGRICOLA, BIOSIS, EMBASE, WPIDS

search terms: ASF/SF2, ASR-2, SR1, arabidopsis, splic? factor#, regulator? element#, express?, promoter#, promotor#